独立発話の繋ぎ合わせによる発話-応答ペアの獲得

赤間 怜奈 †*1　武藤 由依 ‡2　鈴木 潤 †*3　乾 健太郎 †*4
† 東北大学大学院情報科学研究科　‡ 東北大学工学部　* 理化学研究所 AIP センター
{[1]reina.a, [3]jun.suzuki, [4]inui}@ecei.tohoku.ac.jp
[2]yui.muto.p5@dc.tohoku.ac.jp

1　はじめに

自然言語処理分野での深層ニューラルネットワーク技術の発展に伴って，文生成技術も急速な発展を遂げている．特にニューラルネットを用いた機械翻訳（NMT）は，最も研究が盛んに行われている文生成タスクのひとつである．NMT の研究領域では，モデル自体の研究以上に，高品質な対訳データをいかに大量に獲得するかというデータ獲得（あるいは拡張）の方法論の研究が，現在脚光を浴びている [5]．これは，最近の研究成果として，モデルの改良による翻訳品質の向上よりも，適切なデータを獲得し活用する方法論のほうが大幅な翻訳品質の向上に繋がるという研究が報告されているためである [2, 3]．

これと同様のデータ活用の効果が対話応答文生成タスクでも得られるか検討する．対話応答文生成タスクは，ユーザ発話文を入力，システム応答文を出力とすることで，NMT と同一の枠組みでモデル化できる [6, 7]．つまり，機械翻訳と同じタスク構成とみなせる対話応答文生成でも，高品質な発話-応答文ペア集合を獲得できれば，大幅な性能向上が期待できる．しかし，対話応答文生成タスクにおいては，高品質な発話-応答文ペアを獲得する方法論については現状ほとんど議論されていない．

そこで本研究では，対話応答文生成タスクにおいて，NMT モデルを用いることを前提に，その学習データとなる「高品質な」発話-応答ペアを自動で獲得する方法論を議論する．その上で，発話-応答ペアデータの獲得方法として，外部知識（辞書，シソーラス，因果関係データなど）が不要で低コストかつ高速に大量のデータを取得可能な手法を提案する．提案法では，少量の初期データとなる発話-応答ペア集合が存在することを仮定し，ブートストラップ的にデータ拡張を行う．具体的には，初期データから，教師なし単語アラインメント／フレーズ抽出技術を活用し発話-応答ペアのテンプレートを作成する．次に，そのテンプレートに適合する文を大量の発話データから取得し，擬似的に発話-応答ペアとみなすといった方法である．図1に提案法の概略図を示す．

実験では，既存研究 [8] で対話応答生成モデルの学習に用いられた発話-応答ペアデータを用いて，約 1 万の発話-応答ペアを新たに作成した．また，作成した発話-応答ペアに対してクラウドソーシングによる人手評価を行った結果，本手法によってシードデータと同等レベルの品質の高い対話データを獲得できることが確認できた．

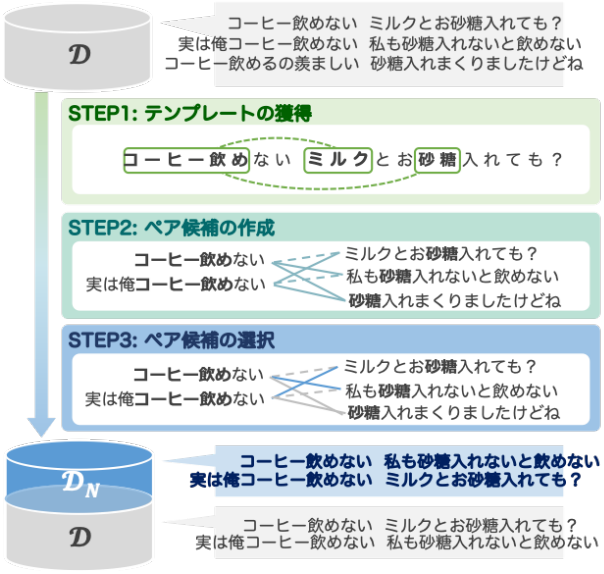

図1　提案法の概観．3 つのステップにより，シードとなる既存のデータ（$\mathcal{D}$, 灰色）のみから新たなデータ（$\mathcal{D}_N$, 青色）を作成し，データを拡張する．

2　独立発話集合からの発話-応答ペアの獲得

2.1　基本のアイディア

$\mathcal{H}$ を独立発話文集合とする．本研究では，$\mathcal{H}$ は比較的容易に大規模に準備できることを仮定する．提案法を簡単に述べると，$\mathcal{H}$ 中から発話-応答ペアとして許容できる発話文ペア (x, y) を自動で獲得する手法である．ただし，$x, y \in \mathcal{H}$ とする．

提案法は，本来，発話-応答ペアとしては成立していなかった未知のペア (x, y) でも，発話-応答ペアとして許容できる組合せが $\mathcal{H}$ 中には一定数存在するであろう，という仮説に基づいている．この仮説は，機械翻訳タスクにおける対訳ペア獲得を考えた場合に，同様の仮説が「成り立たない」とは言えないが，対訳ペアよりも発話-応答ペアの方が，はるかに許容可能な組合せを発見できる確率が高いであろうという直感に基づいている．これは，発話-応答ペアの方が対訳ペアよりも多様性を許容し，一意に定まるものではない，というタスクとしての困難さを逆にうまく利用することで発話-応答ペアの獲得に生かそうという試みである．よって，提案法は対話応答文生成タスクがもつ性質をうまく利用した方法論という位置付けになる．

図1に提案法の概略を示す．提案法では，事前に何らかの方法で構築された，一定の大きさの既存の発話-応

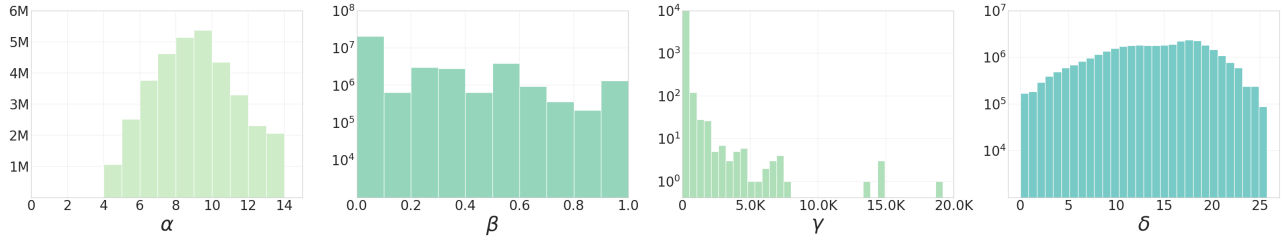

図2　テンプレート選択に用いる各種パラメータの分布．

答ペアの集合 $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{M}$ が存在することを仮定する．ただし，x_i は $\mathcal{D}$ 中の i 番目の発話文，y_i は，x_i に対応する i 番目の応答文とする．また同時に，各 (x, y) は，人間が判断して発話-応答ペアと認定できるデータ，つまり，機械学習の正解データとして利用できる程度の品質が担保されているデータと仮定する．提案法の基本戦略は，既存の発話-応答ペアデータ $\mathcal{D}$ からテンプレートとなるフレーズ対を獲得し，これを手がかりに対応のない独立発話文集合 $\mathcal{H}$ から発話-応答ペアとして成立するペアを獲得することである．提案法は，大きく分けて次の 3 ステップで構成されている．

1. 発話-応答テンプレートの獲得（2.2節）
2. 発話-応答ペア候補の作成（2.3節）
3. 発話-応答ペアの選択（2.4節）

以下に各ステップの詳細を述べる．

2.2　発話-応答テンプレートの獲得

提案法では，発話-応答ペアデータ $\mathcal{D}$ 中に含まれる典型的なフレーズ対 (f, e) を獲得したい．ただし，f は発話文の集合 $\{x_i\}_{i=1}^{M}$ から抽出されたフレーズ，e は応答文の集合 $\{y_i\}_{i=1}^{M}$ から抽出されたフレーズとする．ここでは，得られた典型的なフレーズ対を「発話-応答テンプレート」とよぶ．

まず，文字単位のアラインメントを教師無し学習により獲得し，次に，得られた文字単位アラインメントに基づいてフレーズ対を取得する．この二つの処理は，統計翻訳でのフレーズ抽出の手続きをそのまま活用することを想定する．

次に，得られたフレーズ対に対して，以下の 5 項目の条件を満たすかどうかの判定を行う．ただし，$S(f)$ をフレーズ f に含まれる文字の集合，$\ell(f)$ をフレーズ f の文字数，$c(f, e)$ を構築したフレーズ対の集合におけるフレーズ対 (f, e) の出現頻度，$p(\cdot)$ は出現確率とする．

- f, e の一文字目が記号（「,」「-」など）ではない
- フレーズが一定程度長い：

$$\ell(f) > 1,\ \ell(e) > 1,\ \ell(f) + l(e) > \alpha \tag{1}$$

- f と e が重複しすぎていない：

$$\max\left(\frac{|S(f) \cap S(e)|}{|S(f)|}, \frac{|S(f) \cap S(e)|}{|S(e)|}\right) < \beta \tag{2}$$

- 共起頻度が大きい：

$$c(f, e) > \gamma \tag{3}$$

- 自己相互情報量が大きい：

$$\mathrm{PPMI}(f, e) = \left(\log \frac{p(f, e)}{p(f) \cdot p(e)}\right) > \delta \tag{4}$$

$\alpha, \beta, \gamma, \delta$ は，条件の閾値を決定するハイパーパラメータである．これらの条件を全て満たしたフレーズ対を最終的に発話-応答テンプレートとして採用する．

2.3　発話-応答ペア候補の作成

発話-応答テンプレート (f, e) を手がかりに，対話として成立する可能性のある発話-応答ペア候補を作成する．具体的には，以下の 2 ステップの処理を実行する．

1. 独立発話文集合 $\mathcal{H}$ からフレーズ f を含む発話と e を含む発話をそれぞれ N 個無作為に抽出
2. 組み合わせ総当たりで N^2 の対話ペア候補 $\mathcal{D}_C$ を構築

2.4　発話-応答ペア候補の選択

作成した発話-応答ペア候補 $(u, r) \in \mathcal{D}_C$ のうち，発話 u と応答 r の繋がりを許容できるペアを，最終的な発話-応答ペアに採用したい．発話 u と応答 r からなるペア (u, r) の繋がりの良さ $\mathbf{Assoc}_{\mathrm{s}}(u, r)$ を次のように定義する：

$$\mathbf{Assoc}_{\mathrm{s}}(u, r) := \underset{(e,f)\in(u,r)}{\mathrm{ave}} \left[\mathbf{Assoc}_{\mathrm{p}}(f, e)\right] \tag{5}$$

$$\mathbf{Assoc}_{\mathrm{p}}(f, e) := \lambda \cdot \mathrm{PPMI}(f, e) + (1 - \lambda)(\ell(f) + \ell(e)). \tag{6}$$

ただし $\lambda \in [0, 1]$ はハイパーパラメータ．すなわち $\mathbf{Assoc}_{\mathrm{s}}(u, r)$ は，発話ペア (u, r) に含まれる全ての学習済みフレーズ対[*1] (f, e) について，正の相互情報量 $\mathrm{PPMI}(u, r)$ と文字列長から計算される関連度スコア $\mathbf{Assoc}_{\mathrm{p}}(f, e)$ を計算し，これを平均した値である．$\mathbf{Assoc}_{\mathrm{s}}(u, r)$ が高いほど，発話 u と応答 r の間には統計的に強い関連が認められる，すなわち (u, r) の対話としての繋がりが自然である可能性が高いと考える．

3　実験

3.1　設定

本研究では，初期発話-応答ペア集合 $\mathcal{D}$ として佐藤ら [8] が作成した約 68 万対の対話データを用いた．このデータは，元々は Twitter のリプライチェーンから抽出されたデータである．また，因果関係の知識に基づいてヒューリスティックな方法で抽出されている．データは必ず因果関係があることが前提となっており，かつ最終的に精度重視のフィルタリングを行い，結果として高品質な発話-応答ペアのみで構成されたデータとなっている．

[*1] 2.2節参照．

表1　テンプレートとして獲得したフレーズ対の一例

#	発話側	応答側
1	雨降る	洗濯物干
2	髪切	美容院行
3	歌う	カラオケに行
4	実家に帰	親孝行
5	学校行ってき	お勉強

発話-応答テンプレート獲得（2.2節）の設定：文字単位のアライメントに基づくフレーズ対抽出には IBM モデルベースの統計的機械翻訳ツール GIZA++ [1, 4] を利用した．

発話-応答ペア候補の作成（2.3節）の設定：本研究では，シードとなる発話-応答ペア集合 $\mathcal{D}$ の発話-応答ペアを分解し，発話と応答の全てを独立の発話と考えて発話集合 $\mathcal{H}$ とした．この実験設定は，発話-応答ペアとして存在しているものを一旦解消し適切な繋ぎ変えをおこなうことで，元から存在している発話-応答ペアの他にも発話-応答ペアとして利用できるものを抽出することに相当する．つまり，既存の発話-応答ペア集合 $\mathcal{D}$ 以外のデータを一切使用することなく，データサイズを増大させることを目指す．ただし，提案法では $\mathcal{H}$ の構築方法を $\mathcal{D}$ から構築することに限定しない．本研究では $N = 30$ とし，1 つのテンプレートあたり $N^2 = 900$ の候補を得た．

発話-応答ペアの選択（2.4節）の設定：λ の決定方法としては，ペア候補に含まれる全ての $(u, r) \in \mathcal{D}_C$ を $\mathbf{Assoc}_s(u, r)$ によりランク付けしたものを用いて既存のペアデータの再現能力を平均逆順位（MRR; Mean Reciprocal Rank）により測定し，これが最大となるものを採用する，という方法を用いた．

3.2　結果

発話-応答テンプレートの獲得（2.2節）の結果：文字単位のアライメントの学習により約 4700 万のフレーズ対を得た．得られたフレーズ対のうち，2.2節に示した全ての条件を満たした 148 のフレーズ対をテンプレートとして採用した．表1にテンプレートとして獲得したフレーズ対の例を示す．

また，それぞれの条件に関与するパラメータとフレーズ対全体の分布を図2 に示す．本研究では，$\alpha = 5, \beta = 0.3, \gamma = 14, \delta = 11$ を用いた．

発話-応答ペア候補の作成（2.3節）の結果：$|D_C| = 133,200$ が発話-応答ペア候補として得られた．

発話-応答ペアの選択（2.4節）の結果：図3に示す通り MRR が最大となる λ として，$\lambda = 0.9$（MRR= 0.34）が得られた．

発話-応答ペア候補に含まれる全てのペア $(u, r) \in \mathcal{D}_C$ について $\mathbf{Assoc}_s(u, r)$ を算出した．$\mathbf{Assoc}_s(u, r)$ の分布を図4に示す．本研究では，$\mathbf{Assoc}_s(u, r)$ が全体の上位 5% のペア (u, r) の集合 $\mathcal{D}_{N_5}$ と，上位 10% のペアの集合 $\mathcal{D}_{N_{10}}$ を，それぞれ新たに作成した発話-応答ペアデータとして採用した．それぞれのデータに含まれるペアの数は，$|\mathcal{D}_{N_5}| = 6{,}491$，$|\mathcal{D}_{N_{10}}| = 12{,}981$ である．

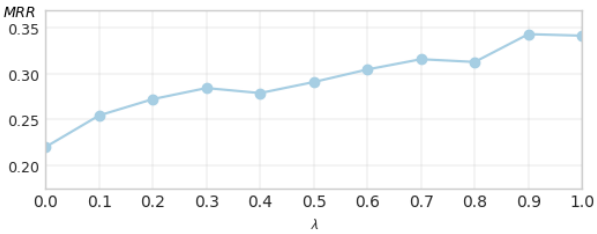

図3　ハイパーパラメータ λ と MRR の関係

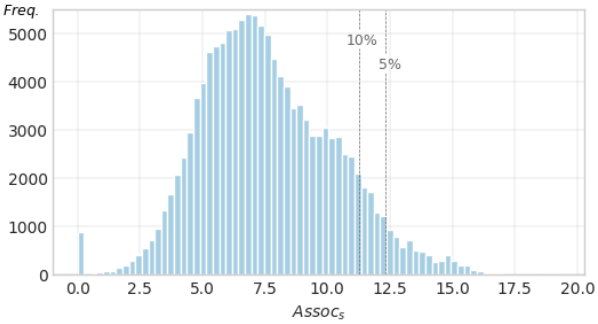

図4　ペア候補の $\mathbf{Assoc}_s$ の分布．$\mathbf{Assoc}_s = 11.26$ 以上が全体の上位 10%，$\mathbf{Assoc}_s = 12.34$ 以上が上位 5% に該当する（図中の破線）．

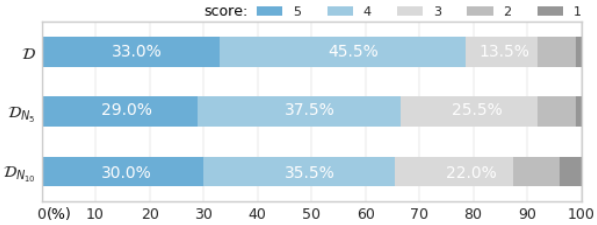

図5　人手によるスコアリング評価の結果．

4　獲得したペアデータの評価・分析

本研究で作成したペアデータが，対話応答文生成タスクの学習データとして適切かどうか，すなわち発話と応答の繋がりが自然な対話となっているかどうかを，人手評価および定性的分析により確かめる．

4.1　人手評価

作成した $\mathcal{D}_{N_{10}}$ および $\mathcal{D}_{N_5}$ について，対話ペアデータとしての質的妥当性を，既存の対話ペアデータ $\mathcal{D}$ との比較により評価する．評価はクラウドソーシングを利用した人手評価によりおこなう．$\mathcal{D}$，$\mathcal{D}_{N_5}$，$\mathcal{D}_{N_{10}}$ からそれぞれ 200 ペアずつを無作為に抽出したものを評価対象とする．ワーカは発話と応答のペアが与えられ，与えられたペアが「対話として許容できるかどうか」を発話と応答の繋がりの良さに着目して 5 段階のスコアリング評価（5: 許容できる, 4: どちらかといえば許容できる, 3: どちらとも言えない, 2: どちらかといえば許容できない, 1: 許容できない）をおこなう．1 つのペアにつき 3 人のワーカが評価をおこない，3 人の平均スコアをそのペアの人手評価スコア $\mathbf{AssocH}_s(u, r)$ として採用する．

人手によるスコアリング評価の結果を図5に示す．

表2　作成した対話ペアと作成過程で算出した $\mathbf{Assoc}_{\mathrm{s}}$ および人手評価による $\mathbf{AssocH}_{\mathrm{s}}$．$\mathbf{Assoc}_{\mathrm{s}}$ の **，* はスコアが全体の上位 5%，10% にそれぞれ該当することを表す．太字はテンプレート部分．

#	発話	応答	$\mathbf{Assoc}_{\mathrm{s}}(u,r)$	human
1	あたりでーす! わー! **コーヒー飲め**ないんです!	ミルクと**砂糖**入れてもダメ?	**13.03	4.67
2	そんなに**視力悪**かったっけ?	悪いよーいつもメガネか**コンタクト**	**13.60	5.00
3	明日**実家に帰**る準備するね	最高の**親孝行**やないか!	**12.55	4.00
4	今日**花粉飛んで**る? 涙と鼻水止まらぬ	俺も**鼻水**半端なかった。風強くて少し寒いね	**15.24	5.00
5	**餃子食べ**たいなあ	京都駅前の**王将行く**ぞ!	*11.51	5.00
6	**宝くじ買う**!	そういうのはまったく**当たら**んでござるよ	*12.32	5.00
7	この時間の時点で**朝起きれる**気しない	ふぁっ? よし**起こして**あげるね	*12.20	4.67
8	わーい**自転車買**うよ	よし、志賀島まで**サイクリング**行こう	*13.18	5.00
9	大変。**腰が痛い**わ。	ありがとー一応**湿布貼っ**て貰ってるよー	**12.57	2.00
10	お風呂**掃除し**て!	あれ**部屋汚**かった?	*11.35	2.67

図5より，作成したペアデータのうち対話としておおむね許容できると判断されたペア（$\mathbf{AssocH}_{\mathrm{s}}(u,r) \in \{4,5\}$ に相当）の割合は作成したペアデータ $\mathcal{D}_{N_5}$ では 66.5%，$\mathcal{D}_{N_{10}}$ では 65.5% という結果であった．既存のペアデータ $\mathcal{D}$ で同様の評価を獲得したペアの割合（78.5%）と比較すると，些か低い値ではあるもののその差は約 10% 程であった．一方，反対に対話として許容し難いと判断されたペア（$\mathbf{AssocH}_{\mathrm{s}}(u,r) \in \{1,2\}$ に相当）に注目すると，その割合は $\mathcal{D}_{N_5}$ では 8.0%，$\mathcal{D}_{N_{10}}$ では 12.5% という結果であった．$\mathcal{D}$ で同様の評価を獲得したペアの割合は 8.0% であり，$\mathcal{D}_{N_5}$ はこれと差がなかった．以上より，提案法は既存のデータとほとんど同等レベルで妥当性のある対話データを作成できることが確かめられた．

4.2　定性的分析

本研究で作成した対話ペアの一例を表2に示す．例示した発話は作成したペアデータ $\mathcal{D}_{N_5}$，$\mathcal{D}_{N_{10}}$ からそれぞれ抽出したもので，人手評価のスコア $\mathbf{AssocH}_{\mathrm{s}}(u,r)$ が高いペアを上部（1-8 行目）に，スコアが低いペアを下部　（9-10 行目）にそれぞれ示す．$\mathbf{AssocH}_{\mathrm{s}}(u,r)$ で高評価の対話ペアについては，全体的に自然な対話が実現されていることがわかる．たとえば 2 行目と 4 行目の対話ペアは発話が問いかけを含んでいるが，応答では問われている内容に対して的確な回答を返している．作成したデータの特徴のひとつとして，発話と応答の間で話のトピックや焦点が共有されているということが挙げられる．これは，統計的な裏付けの取れた特に対話らしいテンプレートを対話の核として使用していることが有効であったと考えられる．一方，低評価の対話ペアについては，関連の強いフレーズ対は含んでいるものの．応答が発話に対するフォローになっていない（9 行目），発話で触れられた対象を対話で正しく扱えていない（10 行目）など，不自然な繋がりが見受けられた．対話としての自然な繋がりをさらに高精度で実現することは，今後の課題である．

5　おわりに

本研究では，対話応答文生成タスクにおいてニューラル翻訳モデルを用いて対話応答文生成を行うと仮定した際に，その学習データとなる発話-応答ペアを自動で獲得する方法論を示した．提案法は，初期データとして利用可能な対応済みの発話-応答ペアデータがあることを仮定した上で，統計翻訳で用いられてきた教師なし単語アラインメント／フレーズ抽出技術を活用し，ブートストラップ的にデータ拡張を行う新しい方法論を提案した．実験では，独立発話集合から効果的に発話を繋ぎ合わせることにより約 1 万の発話-応答ペアを新たに作成した．また，人手評価により初期データと同等レベルの品質の高い発話-応答ペアデータを獲得できることを示した．

今後の発展として，提案法を用いて数百万から数千万規模の発話-応答ペア集合を獲得することを検討している．また，実際にニューラル翻訳モデルの学習データとして利用し，対話応答生成モデルの学習とその性能を評価を行い，データ拡張による対話応答生成タスクの性能向上を実証したい．

謝辞　本研究の一部は東北大学 Step-QI スクールの助成を受けたものです．

参考文献

[1] Peter F. Brown et al. “The Mathematics of Statistical Machine Translation: Parameter Estimation”. In: *Computational Linguistics* 19 (1993), pp. 263–311.
[2] Sergey Edunov et al. “Understanding Back-Translation at Scale”. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2018, pp. 489–500.
[3] Makoto Morishita, Jun Suzuki, and Masaaki Nagata. “NTT’s Neural Machine Translation Systems for WMT 2018”. In: *Proceedings of the Third Conference on Machine Translation: Shared Task Papers (WMT)*. 2018, pp. 461–466.
[4] Franz Josef Och and Hermann Ney. “A Systematic Comparison of Various Statistical Alignment Models”. In: *Computational Linguistics* 29.1 (2003), pp. 19–51.
[5] Rico Sennrich, Barry Haddow, and Alexandra Birch. “Improving Neural Machine Translation Models with Monolingual Data”. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*. 2016, pp. 86–96.
[6] Yuanlong Shao et al. “Generating High-Quality and Informative Conversation Responses with Sequence-to-Sequence Models”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2017, pp. 2210–2219.
[7] Oriol Vinyals and Quoc Le. “A neural conversational model”. In: *International Conference on Machine Learning (ICML) Deep Learning Workshop*. 2015.
[8] 佐藤 祥多 and 乾 健太郎. “因果関係に基づくデータサンプリングを利用した雑談応答学習”. In: 言語処理学会第 *24* 回年次大会. 2018, pp. 1219–1222.